INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
https://doi.org/10.1080/09500693.2018.1521027

Year-long teacher professional development on fifth grade student science outcomes

Carolyn Nichol[a], Alice Chow [a] and Scott Furtwengler [b]

[a]Rice Office of STEM Engagement, Rice University, Houston, TX, USA; [b]Educational Psychology, University of Houston, Houston, TX, USA

ABSTRACT

We first examined the effects of a year-long professional development (PD) programme for elementary science teachers on fifth grade student performance on state-mandated science achievement tests of students from a treatment and a comparison group of teachers in the 2009–2010 academic year. Then, we investigated the longer-term impacts by comparing the 2010–2011 student test results of the teachers one year after receiving treatment in 2009–2010 with the students of teachers who received treatment during 2010–2011. Test scores were analysed using a propensity score matching method to examine the relationship between the PD and student achievement. Results showed that even though the treatment teachers were out of the classroom 20% of the school year to attend the PD, there was no difference between their students' science achievement scores and those of the comparison teachers who were in the classroom every day. This is an important finding because many principals and parents are reluctant to provide teachers with release time for PD. We also determined that students of teachers one year after participating in the PD significantly ($p < 0.001$) with a medium effect size ($\eta^2 = .088$) outperformed students of teachers who had just completed the programme. This suggests that it takes time for teachers to implement new teaching strategies and that to observe the impact of an intervention programme, it may be important to expand the timeframe of the programme evaluation.

ARTICLE HISTORY
Received 2 October 2017
Accepted 4 September 2018

KEYWORDS
Professional development; elementary in-service education; science education; science achievement; and student outcomes

Introduction

We investigate three questions in designing, delivering, and evaluating teacher professional development (PD): 1) How long should the PD last to have an impact on student outcomes? 2) If the PD is held during the school day, does student achievement suffer? 3) When is the best time to evaluate PD outcomes? We have evaluated archival data for two cohorts of teachers that participated in the Rice Elementary Model Science Lab (REMSL), a year-long PD for elementary science teachers focused on inquiry-based, constructivist methods. In a previous study (Diaconu, Radigan, Suskavcevic, and Nichol, 2012), we had examined the teacher outcomes from this intervention. In this paper, we expand upon the prior research to now focus on the student outcomes of

CONTACT Carolyn Nichol cnichol@rice.edu

teachers who participated in the PD. We report on the students' performance on state-mandated fifth grade science exams at the end of their teachers' participation in the programme in 2010 against a comparison group and then compare the PD participants' student test scores the following year with those whose teachers participated that year in 2011. While the teachers in the study are the same, we could not follow the students because in our state there is not a state-mandated science test for sixth grade students.

The REMSL programme had a unique format and design where teachers leave their classrooms one day per week for the entire school year (28 days) to participate in the PD held at an elementary school. Substitute teachers were provided by the school districts while the teachers attend the programme. REMSL was designed and implemented to help elementary teachers in urban school districts acquire the knowledge, skills, confidence, and tools they need to meet the emerging educational challenges in elementary science education. These challenges are much greater for large urban school districts with scarce resources, high percentages of economically disadvantaged and historically underserved minority students, and teachers who lack the necessary preparation in science – the primary population for this form of professional training (Berry, Rasberry, & Williams, 2007). The elementary grades are critical to building science literacy, and yet typically, little to no science is taught at that level (Vasquez, 2005). As Spillane, Halverson, and Diamond (2001) describes, 'Science is largely practiced as a fringe subject, taken up when time allows, but mostly forgotten or treated intermittently and unsystematically' (p. 919). The 2012 National Survey of Science and Mathematics Education (Banilower et al., 2013) shows that only 5% of elementary teachers have degrees in science, engineering, or science education and only 36% of elementary teachers had courses in life science, earth science and physical science. In addition, elementary teachers report feeling less qualified to teach science than any other subject, and that in a typical day, over 30% of K-5 students have no science instruction at all (Fulp, 2002). Elementary teachers face difficult challenges in preparing students for high achievement in science when they lack the necessary content and pedagogical skills to provide students exposure to Science, Technology, Engineering and Mathematics (STEM) fields (Kahle & Kronebusch, 2003; Wenglinsky & Silverstein, 2007).

It is necessary to provide elementary teachers with deep pedagogical content knowledge so that they can engage students in science in the elementary years and expand the pipeline of future scientists and engineers. In a study of scientists and graduate students in science disciplines, it was reported that students' interest in science was most often developed in elementary grades with half as many citing middle school years and even fewer in high school and college combined (Tai, 2008). Thus, exposing and engaging students early in the primary grades to STEM is crucial if we aspire to sustain STEM interest in the later years and cultivate a future workforce in STEM fields (DeJarnette, 2012).

Theoretical framework

While a primary goal of teacher PD is to improve student learning, it is often challenging to establish a direct link between the two because of the many intermediate factors that influence student achievement. Determining the effects of teacher PD on student learning requires understanding a complex system involving adults, who may have misconceptions about science themselves, and students who have widely disparate backgrounds. The

intricate framework of adult learning and student learning is often difficult to analyse. While some studies linking PD with student outcomes have been inconclusive because of data or statistical methodology errors (Wenglinsky, 2000), research exists that indicates a positive relationship between teacher PD intervention and student outcomes (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Heller, Daehler, Wong, Shinohara, & Miratrix, 2012; Johnson & Fargo, 2010; Lumpe, Czerniak, Haney, & Beltyukova, 2012; Polly et al., 2015). Studies have also emerged that report an increase in student achievement observed in the succeeding years after the intervention (Allen et al., 2011; Silverstein, Dubner, Miller, Glied, & Loike, 2009), suggesting that student outcomes should be tracked beyond the initial year. Nevertheless, studies on the longer term effects of teacher PD are still limited and require further investigation.

Literature exists in support of the efficacy of instructional modes in teaching physical science courses for elementary teachers that are built on a constructivist, inquiry-based, and student-centred theory of learning. The findings of empirical studies demonstrating superiority of inquiry-based learning have been widely disseminated among the physics education community (Goldberg, Bendall, Heller, & Poole, 2003; McDermott, Heron, Shaffer, & Stetzer, 2006; McDermott, Shaffer, & Constantinou, 2000). The key features of PD training programmes that promote change in teacher knowledge and teaching practices are commonly regarded as the 'theory of change.' The assumption is that the theory of instruction and teacher change leads to improvement in the quality of science instruction, which in turn has potential to improve student outcomes. As noted by Blamey and Mackenzie (2007), a crucial factor in designing successful reform efforts is ensuring that the programmematic theories of change are clear to teachers and their school administration so that they understand the goals and are aware and invested in the programme outcomes.

The theoretical model proposed by Supovitz and Turner (2000) suggests that effective PD has the potential of producing changes in teaching practices, which in turn translates into higher levels of student achievement. This model proposes that there are six critical components of high quality PD: (1) inquiry and critical thinking, (2) long-term and sustained training (although the number of contact hours is not clear), (3) linkage with ongoing teaching, (4) deepened content knowledge, (5) adherence to high standards, and (6) connection between staff development and school improvement. The researchers also acknowledge the influence of school context variables and state/district policies as powerful mediators in this sequence. The Supovitz model conveys the idea that improvement in student achievement can be attributed to the teacher's implementation of inquiry-based teaching practices and is the basis for the PD in this study. Each of the six components of the Supovitz model are mapped onto the PD model as described below.

Description of the teacher PD program

The REMSL programme began in 2006 with a single partnering school district committed to teacher PD training and broadened to serve 14 districts in 2008. Since then, REMSL has expanded even further to support teachers from 26 school districts. The culminating goal for these teachers was the development of their self-efficacy so that they could apply what they learned in this PD programme to become teacher leaders in their schools (Southerland, Smith, Sowell, & Kittleson, 2007). These teacher leaders could then enrich their

communities through training and mentoring sessions, sharing lesson plans developed through the programme, leading science fair and school-wide science nights, and writing grant applications to strengthen their schools and propagate the effects of their training.

The REMSL PD programme was designed to incorporate the six features of highly effective teacher training programmes in science, as described earlier by Supovitz. First, the programme is based on constructivism, which involves building student knowledge through interactions with real 'physical and/or social' experiences and the synthesis of the knowledge gained with prior knowledge (Liang & Gabel, 2005). Moving to the context of the classroom, social constructivism highlights the way small groups of students negotiate as they contribute their understanding of scientific phenomena to group knowledge building (Southerland et al., 2007). Following the guidelines of the National Research Council (2000), learners in the classroom engage in inquiry science and develop critical thinking by asking scientific questions, investigate evidence that evaluates these questions, and use this evidence to develop and justify explanations. Tai, Loehr, and Sadler (2005) have reported on the positive effects of inquiry science in urban settings. Because teachers are actively learning about active learning, the REMSL programme models teaching practices demonstrated to be effective (Armour & Yelling, 2004). The teacher training in science in the REMSL programme is implemented in the 5E inquiry model for lesson delivery (Bybee et al., 2006) and is additionally supported by technology-based science curriculum including simulations, graphing, data analysis, computer-based manipulative, non-linguistic representations, and progress assessments that are rooted in exemplary content pedagogical methods.

Second, REMSL was designed to provide long-term, sustained training. This construction aimed to bring teachers to an elementary school for a full day of training each week for an entire academic year and was based on the assumption that a well-conceptualized PD training programme has a higher likelihood of success if it is extensive and embedded in the regular work week of teachers (Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009; Wayne, Yoon, Zhu, Cronen, & Garet, 2008; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). This delivery mode is intended to preserve the fidelity of implementation (FOI), defined as the estimation of how well an intervention is implemented according to its original programme design (Dusenbury, Brannigan, Falco, & Hansen, 2003; Mowbray, Holter, Teague, & Bybee, 2003). Effective professional programmes are characterised by extended duration (Desimone, 2009; Wayne et al., 2008). In the meta-analysis of studies reviewed by Yoon et al. (2007), evidence suggests that PD is more likely to be 'effective' if delivered in larger doses. In the case of REMSL, the project participants receive 196 contact hours of science content and pedagogy training per year resulting in both a large 'dose' and large duration.

Third, because the REMSL PD is aligned with the districts' scope and sequence, or the recommended teaching order for elementary science, the PD occurs in close time proximity with its implementation in elementary science classrooms. As a result, the PD is strongly linked with ongoing teaching. As previously described, the FOI was evaluated for participants in the REMSL programme through independent observers using validated instruments reported in a multi-year study about the teacher outcomes of the PD programme (Diaconu et al. 2011). These observers were not affiliated with the programme and had extensive experience as classroom science teachers or science specialists. They

were trained on how to use the observation protocol by the university staff and were "blind" to whether they were observing a PD participant or a comparison teacher's classroom.

Fourth, the professional training is focused on improving teacher *knowledge of science content* which has been found to be an essential ingredient in a programme's effectiveness (Hill et al., 2008; Kennedy, 1998; Supovitz & Turner, 2000). The Kennedy study assessed a large number of PD programmes by the level of subject matter provided to teachers. On the basis of the analysis of effect sizes, she concluded that training focused on developing teachers' subject content knowledge demonstrated the greatest influence on student learning. Kennedy's meta-analysis instigated other research groups to test the same research hypotheses (Blank & De la Alas, 2009; Desimone, Porter, Garet, Yoon, & Birman, 2002; Kahle & Kronebusch, 2003; Wenglinsky & Silverstein, 2007), and reach similar conclusions.

The fifth component of the Supovitz theoretical model links the PD to high standards. Because REMSL is intended for teachers in Texas, the PD was tightly aligned to the Texas standards known as the Texas Essential Knowledge and Skills (TEKS). The 1997 version of the TEKS was a reform effort designed to reflect the standards-based recommendations of the National Research Council and the American Association for the Advancement of Science. According to a national evaluation of science standards (Mead & Mates, 2009), Texas standards are considered 'generally comprehensive except for creationist jargon.' In addition, Moore (2001) shared a similar assessment of high school biology TEKS.

The sixth characteristic of highly effective PD programmes is that they are school-based, connected to school improvement, and integrated into the regular work week of teachers (Hawley & Valli, 1998; Joyce & Showers, 2002). The REMSL programme is offered during the school day and has a campus support component, establishing one-on-one training to programme participants on their home campuses. In addition, school principals are invested in the programme and provide substitute teachers and facilitate campus-wide implementation such as providing time for REMSL participants to share their knowledge in Professional Learning Communities (PLCs) or in regularly scheduled grade level meetings.

Although the programme design was influenced by the literature on effective PD (Desimone, 2009; Wayne et al., 2008) and reinforced by current studies (Desimone, 2011), three aspects of the in-service teacher training programme remain unique to the REMSL programme:

(1) Comprehensiveness of the intervention: providing teacher training in science content and pedagogy, teacher support through science materials and resources, and a one-on-one campus support component.
(2) Mode of instructional delivery: offering one full day (seven hours) each week throughout the academic year.
(3) Intensity of the intervention, which is based on the university model and exceeds all 'direct contact hour' standards described in the literature: about 80 teachers are trained weekly for 196 h per academic year.

These three elements embodying the core of the REMSL programme are clarified below.

Comprehensiveness of the intervention

Content

The REMSL science content includes *Nature of Science* (designing and implementing investigations, observing and measuring, analysing and interpreting graphs, and drawing inferences using models); *Life Science* (life cycles, inherited versus learned characteristics, adaptations, food webs, etc.); *Earth Science* (weather and atmosphere, cycles of the Earth, soil properties, natural resources, changes to land, solar system, etc.); and *Physical Science* (properties of matter, mixtures and solutions, boiling and melting points, energy and light, electricity and sound, force and motion, etc.). These major topical areas are tightly aligned with the state standards and presented to the participants through hands-on, inquiry-based activities. The science activities are flexible enough to be either directly implemented or modified by teachers and then utilised in elementary classrooms.

Pedagogy

The REMSL programme emphasises the use of research-based teaching strategies, including student engagement activities, quality questioning techniques, vocabulary in context, science and literacy integration, and other methods of intervention that have been tested and proven to be effective in increasing student learning. These methods are successful in a wide range of educational settings and across student populations, including students with limited English proficiency and those that are economically disadvantaged. These populations comprise a large proportion of the student sample involved in the study.

Teacher support

To maximise the impact of lessons learned during PD sessions, teachers trained through REMSL are supported with materials and resources for effective implementation of learned practices in their elementary classrooms. Teachers receive items ranging from basic lab supplies, such as balances and graduated cylinders, to professional textbooks. Additionally, REMSL offers a complimentary online curriculum which is based on the 5E model (Bybee et al., 2006) and aligned with the state standards in science. The curriculum serves as a teacher resource in science, offers interactive activities for students, and contains 'reading passages' and 'math connections', therefore bridging curriculum across subjects which is highly appropriate for the context of elementary education.

Participants are also offered one-on-one campus support. Teachers receive important feedback from REMSL staff on their teaching practices and get exposure to other critical activities, including co-teaching and receiving support in presenting relevant science content pedagogy to other peer-teachers from their campus or district. Another feature of effective programmes for in-service teachers is being school-based and integrated into the daily work of teachers (Hawley & Valli, 1998). This 'one-on-one' form of interaction between the instructional team member and the teacher is among the most expensive approaches to PD available, but empirically shown to be the most effective (Joyce & Showers, 2002).

PLCs support the transition from direct instruction to inquiry practices (Armour & Yelling, 2004). Working in PLCs, teachers-as-students participated in science learning experiences, studied constructivist pedagogical practices for diverse learners, and received

leadership training. Using reflective journals, teachers were guided to define inquiry, the ways they were incorporating inquiry into their teaching, and how their conception of scientific inquiry was changing through their classroom practice (Moseley & Ramsey, 2008). Finally, through their digital portfolios, teachers could trace the journey of their science content learning and changing teacher practices throughout their year of implementation in their own classrooms.

Mode of instructional delivery

The mode of instructional delivery, one full day of training each week for an entire academic year, was based on the assumption that a well-conceptualized PD training programme has a higher likelihood of success if it is extensive and embedded in the regular work week of teachers (Wei et al., 2009; Wayne et al., 2008; Yoon et al., 2007). This delivery mode is intended to preserve the FOI. The REMSL intervention occurs in close time proximity with its implementation in elementary science classrooms, and the FOI was evaluated for participants in the REMSL programme through independent observers using validated instruments as reported previously in a multi-year study about the teacher outcomes of the PD programme (Diaconu et al. 2011).

Intensity of the intervention

With seven hours of PD per session, REMSL's highly intensive intervention provided teachers with 196 h of PD, much longer than the average length of 35 h as reported in Garet's large study (2001), for example. Four cohorts of participants (about 20 teachers per cohort) attend class at the model science lab held at an elementary school one day per week for 28 weeks throughout the academic year. Daily activities are divided into content-focused morning sessions (four hours) and pedagogy-centred afternoons (three hours). During the morning sessions, participants engage in inquiry-based science lessons and conduct scientific investigations which meet the TEKS state standard for education content. The afternoon sessions are focused on the utilisation of effective teaching practices and construction of lessons for their students. Together, these sessions offer a variety of innovative pedagogical techniques needed for successful teaching and learning of science.

With supplies and materials provided by the REMSL programme, teachers are better equipped to transfer what they learned in the lab each week and implement it in their classrooms. The yearlong, weekly format of the PD programme allows teachers to discuss their classroom experiences with the new lessons soon after they had learned them and implemented them and with sufficient time to reinforce what they learned. Electronic portfolios are used to document the teachers' pedagogical growth, science content mastery, leadership growth and changes in attitudes toward science. Videos of teaching experiences, training team visits to teachers' classrooms, and portfolio evidence facilitate the development of teacher progress. An online resource was developed for the distribution and sharing of files and the building of a library of web-based resources.

Overview of the research methods

Data was collected from the 2009–2010 to 2010–2011 school years. In the 2009–2010 data, we examined differences in science achievement between students whose teachers were in their first year of the PD (treatment) and students whose teachers were not participating in the PD (comparison). In the 2010–2011 data, we investigated differences in science achievement between students whose teachers were in the year following their PD (catalyst) and students whose teachers who were in their first year of the PD (treatment). Although ideally these data would be considered in a single model with three groups (comparison, treatment, catalyst), this was not possible given that some teachers were included in the treatment group in 2009–2010 and also in the catalyst group in 2010–2011. Furthermore, we considered conducting a repeated measure comparison looking at within-teacher change from the treatment year to catalyst year of the PD. However, given that an additional year of PD is confounded with an additional year of teaching experience, we opted to constrain our analyses to between-group differences.

A quasi-experimental research design was utilised to assess the effects of a yearlong PD teacher training in elementary science (REMSL) on student academic achievement in science. Specifically, the effects of the programme on the student performance of participants and comparison teachers on the Texas Assessment Knowledge and Skills (TAKS) science test, which was developed by the Texas Education Agency, to meet the requirements of the No Child Left Behind policy. Tests in reading, writing, math, science, and social studies were mandated in designated years, which included an assessment of elementary science knowledge for fifth grade students. The TAKS science scale ranged from 910 to 2800 where scores above 2100 are considered passing. In our analyses, these test scores were treated as a continuous measure. It should be noted that fifth grade is the only elementary grade with a state mandated science test. While mathematics is tested in third, fourth, and fifth grades, there is no TAKS science data for third or fourth grade nor is there data for laboratory teachers or science specialists who sometimes participate in our programme but who are not the teacher of record. The high stakes TAKS test was factored into teacher performance pay and administered once a year during the month of March and remained in place until the 2011–2012 academic year, when it was replaced by the State of Texas Assessments of Academic Readiness (STAAR) test (Texas Education Agency, 2013).

The TAKS data were provided to REMSL staff researchers from participating districts. As part of their REMSL application process, teachers submitted a signed Principal Agreement letter stating that s/he could be released from class to attend the programme during the school day; that the school would provide a substitute or other coverage for the teacher (one day/week); and that the principal would facilitate the collection of student TAKS scores and demographic information, pending district approval. In addition, the superintendents of all school districts whose teachers participated in the programme agreed to provide the programme researchers with student level data prior to the admission of teachers into the programme. The districts' research and accountability offices provided the programme staff with the student level data as per Institutional Review Board approved protocols.

The target population for this programme was third, fourth, and fifth grade science teachers, their students and school principals from Region 4 Houston area districts where 56% of students were economically disadvantaged, 48.6% at-risk, and 20% English

language learners (Texas Education Agency-Lonestar, 2013). Moreover, we included teachers who were themselves members of historically underserved minorities as project participants. We also wanted to obtain a representative sample of participants that would reflect the population of students and teachers from the Houston area. We sought to train teachers who were in need of PD in elementary science teaching and possessed the potential to become leaders in their campuses and districts.

Teachers that entered the 2009–2010 applicant pool and selected based on the aforementioned areas of emphases were randomly assigned either to the Treatment or Comparison group. However, teachers that participated in the comparison group the previous year (2008–2009) were automatically invited to participate in the REMSL programme the year after (2009–2010) as part of the protocol design. Thus, the designation of teachers for the treatment and comparison groups was not completely randomised. The single teacher per school selection criteria was enacted after discussions with school districts as a way to minimise costs since schools or their districts paid for substitute teachers and to ensure that the PD programme served the largest number of schools. If the teacher was assigned to the treatment group, the teacher would participate in all the activities of the REMSL training. On the other hand, if the teacher was assigned to the comparison group, then the teacher would continue with their regular teaching assignment without participation in the REMSL training.

Teacher demographic information provided by Region 4 ESC (Texas Education Agency, 2011) which includes the greater Houston geographic area was compared with the REMSL teacher profile to examine how closely the participants represented the local teacher population. Region 4 data indicated that the majority of teachers in the region were 78% female, 19% African American, 17% Hispanic, and 59% Caucasian. The majority of teachers who participated in the 2009–2010 REMSL training had similar characteristics to the Region 4 demographic profile. For example, 79% of participants were female, 7% African-American, 18% Hispanic, 64% Caucasian and 12% (other).

For the present study, the analysis the researchers initially utilised was a multiple linear model, also known as a hierarchical linear model (HLM). The model resulted, however, in a violation of the assumptions of 1) normal distributions of variables and of 2) homoscedasticity. Variables that are not normally distributed in a regression model may result in misinterpretations of significance tests and relationships. Further, multiple regression requires that the variance of error terms must be similar (i.e. homoscedastic) across the values of the independent variables. Heteroscedastic data may increase the possibility of Type I error (Osborne & Waters, 2002). Since the data were determined to be non-normalized, log-transformation of non-normalized data was attempted without success. Therefore, HLM was deemed less suitable due to its sensitivity to small group size and missing or insufficient data causing attrition of the data that remained useable. Studies have used propensity score methods (Furtwengler, 2015; Van Overschelde, 2013) which can better achieve an unbiased estimate of the treatment effects by exploiting the available data for optimal matching of the covariates of the treatment and comparison groups. In this study, propensity score matching served to account for differences in baseline characteristics between the treatment and comparison groups in our quasi-experimental research design. The propensity score was estimated using a logistic regression model in which the treatment status is regressed on observed baseline characteristics and covariates balanced between the treatment and comparison groups. Among cases who share similar propensity

scores in the present study, the distribution of the observed baseline characteristics were the same between the two groups, controlling for a greater number of variables and producing an unbiased estimate of effect for the intervention that could not have been obtained by comparing outcome measures between the two groups.

The study explored the following research questions:

(1) Do students of the teachers in the comparison group differ from the students of the teachers in their first year of the intervention on a science achievement test?
(2) Do students of the teachers in their first year of the intervention differ from the students of the teachers a year after the intervention on a science achievement test?

To address these questions, fifth grade TAKS student data from the 2009–2010 to 2010–2011 school years were examined. Utilising a propensity score matching method, the study assessed whether there was a significant relationship between a teachers' participation in the REMSL programme and student TAKS science scores. Analyses with covariate adjustment using propensity scores have been effectively utilised in educational studies where selection and assignment into groups is not random and not based on clear selection criteria (Furtwengler, 2015). Propensity scores were calculated as a method for better estimation of treatment effect on the criterion variable of TAKS scores between the groups (comparison and treatment (first year); treatment and catalyst (second year)). The propensity score is defined as the probability of receiving treatment based on measured covariates (Thoemmes & Kim, 2011): $e(x) = P(Z = 1 \mid X)$, where $e(x)$ is the propensity score, P is the probability, $Z = 1$ indicates receipt of treatment with values 0 for control or comparison group and 1 for treatment group, and X represents the observable characteristics. A propensity score for each student to determine predicted probability for enrolment in courses taught by REMSL participants was calculated using five observable variables: gender, ethnicity, economic status, limited English proficient status (LEP), and TAKS math scale score. In essence, calculating and utilising a propensity score controlled for these five observable characteristics or covariates (D'Agostino, 1998). The resulting propensity score was then used to match pairs between the two groups providing optimal balance. Optimal balance, however, does not imply perfect balance. As D'Agostino (1998) notes, "Although the idea of finding matches seems straightforward, it is often difficult to find subjects who are similar (that is, can be matched) on all important covariates, even when there are only a few background covariates of interest (p. 2268). Indeed, in the present study, although the comparison and treatment (Catalyst) groups differ on the observable characteristics of economic disadvantage and ethnicity, the propensity model reduces bias based on all five observable characteristics, providing a better estimation of effect than an analysis of variance between the treatment and comparison groups without matching on the covariates. In other words, all five observable characteristics are accounted for in the model, providing a more accurate analysis of the outcome variable.

A binary logistic regression was used to estimate the propensity scores because the dichotomous assignment to either the comparison or treatment group served as the outcome variable and the selected observable variables were the predictors. Analyses of Variance (ANOVAs) were conducted to estimate the average treatment effect (ATE) in each cohort. For the present study, one-way ANOVAs were the appropriate statistical analyses for three reasons: (1) the observable characteristics that will be used as the matching

Table 1. 2009–2010 Descriptives of matched groups ($N = 876$).

Category	Comparison group	Treatment group
Gender		
Female	233	226
Male	205	212
Ethnicity		
American Indian or Alaskan Native	1	1
Asian or Pacific Islander	53	54
African American	65	84
Hispanic	153	139
White, not of Hispanic origin	166	160
Economic Code		
Not identified as economically disadvantaged	292	262
Eligible for free meals	146	175
Eligible for reduced-priced meals	0	1
LEP Status		
Non-LEP	379	398
LEP	59	40
TAKS math scale mean scores	721.25 (103.44)	711.13 (89.47)

variable will be balanced as a result of the propensity calculation procedure; (2) the nature of the data collection is archival; and (3) the research design is quasi-experimental.

Valid propensity scores for the 2009–2010 and 2010–2011 cohorts were generated for 1591 (253 missing) and 2042 cases respectively. Case matching between the two groups resulted in 437 matched pairs ($N = 876$) in the 2009–2010 cohort and 492 ($N = 988$) matched pairs in the 2010–2011 cohort (Tables 1 and 2).

Results and discussion

For the 2009–2010 cohort (Table 3), the difference on students' TAKS science scores between the students of the REMSL programme teachers (2009–2010 REMSL

Table 2. 2010–2011 Descriptives of matched groups ($N = 988$).

Category	Treatment group	Catalyst group
Gender		
Not Reported	210	25
Female	130	247
Male	154	222
Ethnicity		
Not Reported	264	100
American Indian or Alaskan Native	2	45
Asian or Pacific Islander	0	5
Hispanic	166	150
White, not of Hispanic origin	62	194
Economic Code		
Not identified as economically disadvantaged	256	334
Eligible for free meals	231	144
Eligible for reduced-priced meals	7	16
LEP Status		
Non-LEP	425	446
LEP	69	48
TAKS math scale mean scores	620.16 (129.48)	745.59 (119.05)

Table 3. 2009–2010 Group mean scores, standard deviations, and *P*-value on 2009–2010 TAKS science scale.

	N	Mean	Standard Deviation	*P*-value
Comparison	438	2371.84	270.99	
Treatment	436	2360.67	249.87	$p > 0.05$
Total	874	2366.26	260.58	

participation) and comparison teachers was not statistically significant, $F\,(1, 872) = 0.40$, $p > .05$. For the 2010–2011 cohort (Table 4), however, there was a significant difference on the TAKS scores between students of the REMSL teachers (2010–2011 REMSL participation) and REMSL catalyst teachers (2009–2010 REMSL participation), F (1, 982) = 95.34, $p < .001$, $\eta^2 = .088$. Teacher participation at the catalyst level accounted for 8.8% of the variance on mean TAKS science scale scores.

When designing this programme, there had been some concerns expressed that a PD programme that takes an elementary teacher out of the classroom one day per week for an entire school year would have negative effects on student achievement during that year. However, results from the 2009–2010 data (Table 3), the TAKS scores of students of the teachers in the comparison group did not differ from the students of teachers in the first year of the intervention (treatment). Because the programme was designed to align with the scope and sequence of the largest school districts the programme served, teachers reported that they were implementing the PD content during the intervention year. Published results (Diaconu et al., 2012) also support that most teachers were trying to use the content and pedagogy that was provided through the programme. However, challenges with implementation included schools' resistance to change long-standing educational approaches. PD literature also suggests that significant modifications in pedagogy are more likely to occur in the years following training as teachers adapt and incorporate the new methods into their practices (Wayne et al., 2008). Therefore, our results that showed improved science achievement for students of teachers in their second year of programme implementation is promising and indicative of some success.

The results from the 2010–2011 data (Table 4) show that the students of catalyst teachers outperform students of treatment teachers with catalyst teacher participation accounting for 8.8% of the variance on mean TAKS science scores, corresponding to a medium effect size (Cohen, 1988). One explanation is that the teachers are back in the classroom full time in the year after the REMSL programme participation to better implement the intervention especially since the PD programme removed participating teachers from the classroom one out of five days of the week during that time. Another possible reason for the increase in student test scores is that while teachers may have had time to try to implement some components of the PD when they were participating in the programme, they are able to more fully apply programme practices after the completion of the entire programme. Change is a process and it takes time for teachers to use new curricular materials and implement new teaching strategies (Anderson, 1998). Barriers exist to implementation including time to explore, learn, and discuss changes in their teaching (Loucks-Horsley, Hewson, Love, & Stiles, 1998). The PD intervention incorporates long-term, coherent PD that includes time for exploration and experimentation, time for collaboration with programme staff and teachers at their local schools, and time for participants and staff to work together to try new teaching methods, hone their teaching

Table 4. 2010–2011 Group mean scores, standard deviations, *P*-value, and effect size on 2010–2011 TAKS science scale.

	N	Mean	Standard Deviation	P-value	η^2
Treatment	491	2232.15	224.81		
Catalyst	493	2385.19	265.10	$p < 0.001$	0.088
Total	984	2308.83	257.35		

approaches, and share these practices after they had used them in their classes. Our observations and interviews from our previous study (Diaconu et al., 2012) reveal that teachers are learning and excited about the constructivist teaching methods during the programme, but the impact on the students is more evident after they have ultimately completed the programme.

Researchers have shown that teachers tend to be conservative and averse to changing their own practice (van Driel, Beijaard, & Verloop, 2001) and rather than changing it dramatically, they build on their practices in a gradual fashion (Bereiter & Scardamalia, 1993). For example, Allen et al. (2011) observed student achievement gains that occurred the year after teacher intervention while others have shown that it may take years for PD interventions to impact student test scores. Silverstein found that students of teachers who participated in research experiences three and four years earlier had higher passing rates on the New York Regents exam (Silverstein et al., 2009). Supovitz (2001) also found that it takes several years to translate PD experiences into practice. However, the aforementioned studies documenting latent student achievement gains are among the few. Allen et al reports only two rigorous studies linking teacher professional development and student achievement and both were in mathematics. Silverstein's research focuses on science but at the high school level and reports gains after three to four years. This paper presents a rare examination of elementary science outcomes as impacted by a unique teacher professional development programme that has been demonstrated to be impactful on teachers after one year. Furthermore, Texas mandates the TAKS science exam for only fifth and eighth grades. Thus, student science outcomes are difficult to capture and linking to teacher intervention even more challenging. Our research was feasible due to our long-running collaborations with Houston school districts who provided the student data and who have committed their teachers to our programmes due to its established success.

Conclusion

The results of propensity score analyses indicated that fifth grade science students whose teachers who received PD that year did not differ from a comparison group based on TAKS science scale scores. However, the students of catalyst teachers who received PD one year prior scored significantly higher on the same measure than the students of teachers who received PD that year with the catalyst teacher participation accounting for 8.8% of the variance, a medium effect size. These findings suggest that the teachers may be trying to implement their newly acquired content and pedagogy while they are participating in the year long programme, but it is not until the following year that they can fully utilise their newly developed skills. It takes time to implement new teaching strategies and evaluating the programme a year after teachers' participation can reveal a significant increase in science scale scores.

Exposure to an intensive job-embedded teacher training programme in science, supported by curriculum which is tightly aligned with state recommended standards and based on science content through inquiry, resulted in a significant improvement of student achievement test scores in the year following the PD. The programme removed participating teachers from the classroom one out of five days of the week during the year of programme intervention. Despite the teachers' 20% absence during the school year, the science achievement test scores did not suffer as no significant difference between the treatment students and comparison students was detected in Year 1. The improvement shown in the Year 2 data reflects the more fully developed implementation of learning received in Year 1.

Most importantly, these results reveal to providers of PD that they must be cognisant of the time that it takes for teachers to implement new content and to change their teaching practices. Rather than assessing outcomes only at the end of the programme, it is important to continue to follow the teachers and their students after they have completed the programme and can fully implement the content and pedagogy in the subsequent school years. During the year following this long term, intensive PD, teachers will have had time to exercise and cultivate their skills as reflected in the results of improved student outcomes. In essence, Year 1 consists of the delivery of the PD programme while Year 2 begins the true implementation of these learned practices. A strong statement can be made about the REMSL programme design where students' standardised test scores are significantly impacted just one year after participation rather than several years later in other studies (Silverstein et al., 2009; Supovitz, 2001). In addition, this study reports on elementary students' science achievement whereas more literature focuses on secondary outcomes and in mathematics. These analyses will be useful to the education community regarding the eventual student achievement gain from participating in a long term PD programme with no apparent detrimental impact the year the teacher is pulled from class one day a week to participate in the PD and provide insights into the longer term effects of teacher PD and when to evaluate PD outcomes.

Study limitations

Improvements in research design could be achieved in a few ways, namely controlling the breadth of the programme and limiting enrolment to a specific school district. Confining a study to one school district could produce a better comparison study with well-matched demographics and profiles between groups providing a more rigorous research study. Although this would produce research data that is more specific to one type of district but lead to more accurate statistical controls, the primary goal of the REMSL programme is to improve student achievement and engagement in science in the greater Houston region, not confined to one school district. Our strong partnerships with the many school districts in the Houston area has granted the large amount of data received from 17 school districts; however, not all of the data was usable. Some districts submitted incomplete information regarding teacher identity or student demographics which had to be excluded for the original hierarchical linear modelling analyses. As a result of the limited data in some racial categories, this particular demographic was collapsed into white and non-white. The usable data from the treatment group was also decreased to correspond to the size of the comparison group. Improved collaboration and agreements with

school administration upfront may help lead to more uniform and comprehensive data collection. Finally, cultivating stronger partnerships with schools may enhance the support teachers receive in their efforts to adopt innovative and transformative teaching practices.

Another issue that arose in the study is teacher-teacher interactions. Part of the funding for the PD required that the teachers in the programme provide mentoring to colleagues at their campuses. While this mentoring was beneficial to the schools, it could have led to some 'contamination' of the study. While only one teacher per school participated in the programme in a given year, teachers would share lesson plans and programme information with other teachers at their schools. Since schools with teachers participating in the PD would be more inclined to apply to the programme in subsequent years, especially if they wanted to build their science teaching capacity in their school, it is quite possible that the comparison teachers may not have not been naive to the programme materials. In an ideal study, it would be best to recruit from separate schools each year. However, in the ideal PD programme, it would be best to include all teachers on a campus because this promotes collaboration and has been shown to support implementation and adoption (Hord & Roussin, 2013).

While the study controlled for important student variables like student socioeconomic background, it did not control for school environment, curriculum, tutoring programmes or other student factors. Furthermore, this study did not control for teacher variables including teaching experience and academic background, which could clearly interact with the effectiveness of the PD in producing positive student effects.

Because of the intensity of the REMSL programme, the teachers who participate are highly motivated and thus, self-selection biases are possible. However, we did attempt to control for this bias by using teachers who had applied for the programme as our comparison group. Moreover, teachers who participated in the programme had the support of their school administrators who paid for substitute teachers while teachers were out of the classroom. Often principals would prefer to send their most effective teachers to the programme so that they would come back and share the programme materials with the rest of the school. In addition, the substitute teacher was also an uncontrolled factor. Lesson plans were provided for teachers to give to substitute teachers during their absences, but the effectiveness of the substitute teachers were not part of the research study.

Future research

The overall promising results inspire further investigation into other areas of study. Classroom configuration, such as lab teacher or science coach, as a fixed effect can apprise of additional potential influences in teaching efficacy and test results. Another area of exploration includes the effective outcomes from the PD programme on students' math and reading scores. Plans are underway to evaluate multi-year data with comparison groups correlated with various survey results to determine the impact of teacher content knowledge and teaching efficacy on student achievement in larger studies. The availability of centralised data at research centres and data clearinghouses will mitigate the challenges with data collection and expand the scope of research studies. In conclusion, further research is recommended to follow teachers who participate in long-term PD similar to the REMSL programme and explore the impact on student achievement as

teachers continue to implement the programme in the succeeding years. The results reported in this paper suggest that teachers continue to grow after completing the programme and become more student-centric in their practices, which is expected to translate to greater student achievement in science.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Texas Education Agency [grant number 106944047110001] and a gift from ConocoPhillips.

ORCID

Alice Chow http://orcid.org/0000-0002-3703-4453
Scott Furtwengler http://orcid.org/0000-0002-5904-1611

References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333* (6045), 1034–1037. doi:10.1126/science.1207998

Anderson, J. (1998). *Service-learning and teacher education*. Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education. [ED421481].

Armour, K. M., & Yelling, M. (2004). Professional development and professional learning: Bridging the gap for experienced physical education teachers. *European Physical Education Review, 10*, 71–93.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago and La Salle, Illinois: Open Court Publishing Company.

Berry, B., Rasberry, M., & Williams, A. (2007). *Recruiting and retaining quality teachers for high-needs schools: Insights from NBCT summits and other policy initiatives, national strategy forum.*

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation SAGE publications (Los Angeles, London. *New Delhi and Singapore, 13*(4), 439–455. doi:10.1177/1356389007082129

Blank, R. K., & De la Alas, N. (2009). *Effects of teacher professional development on gains in student achievement: How meta-analysis provides scientific evidence useful to education leaders.* Washington, DC: Council of Chief State School Officers.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, H., & Landes, N. (2006). *The BSCS 5E instructional model: Origins, effectiveness, and applications*. Colorado Springs: BCSC.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillside, NJ: Erlbaum.

D'Agostino, R. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17*(19), 2265–2281.

DeJarnette, N. (2012). America"s children: Providing early exposure to STEM (science, technology, engineering and math) initiatives. *Education, 133*(1), 77–84.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures'. *Educational Researcher, 38*(3), 181–199.

Desimone, L. M. (2011). R and D: A primer on effective professional development. *Phi Delta Kappan, 92*(6), 68–71.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81–112.

Diaconu, Dana Viorica, Radigan, Judy, Suskavcevic, Milijana, & Nichol, Carolyn. (2012). A Multi-Year Study of the Impact of the Rice Model Teacher Professional Development on Elementary Science Teachers. *International Journal of Science Education, 34*(6), 855–877. doi:10.1080/09500693.2011.642019

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237–256.

Fulp, S. L. (2002). *2000 national survey of science and mathematics education: Status of elementary school science teaching*. Chapel Hill: Horizon Research. Retrieved from http://www.horizon-research.com

Furtwengler, S. (2015). Effects of participation in a post-secondary honors program with covariate adjustment using propensity score. *Journal of Advanced Academics, 26*, 274–293.

Goldberg, F., Bendall, S., Heller, P., & Poole, R. (2003). *Constructing ideas in physical science: Field test version*. Armonk, NY: It's About Time.

Hawley, W. D., & Valli, L. (1998). The essentials of effective professional development: A new consensus. In L. S. Darling Hammond, & G. Sykes (Eds.), *The heart of the matter: Teaching as a learning profession* (pp. 86–124). San Francisco: Jossey-Bass.

Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching, 49*(3), 333–362.

Hill, C. H., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511.

Hord, S., & Roussin, J. (2013). *Implementing change through learning; concerns-based concepts, tools, and strategies for guiding change*. Reference and Research Book News.

Johnson, C. C., & Fargo, J. D. (2010). Urban school reform enabled by transformative professional development: Impact on teacher change and student learning of science. *Urban Education, 45*(1), 4–29.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kahle, J. B., & Kronebusch, M. (2003). Science teacher education: From a fractured system to a seamless continuum. *Review of Policy Research, 20*(4), 585–602.

Kennedy, M. (1998). *Form and substance in inservice teacher education*. Madison, WI: National Institute for Science Education, University of Wisconsin-Madison.

Liang, L., & Gabel, D. (2005). Effectiveness of a constructivist approach to science instruction for prospective elementary teachers. *International Journal of Science Education, 27*(10), 1143–1162.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

Lumpe, A., Czerniak, C., Haney, J., & Beltyukova, S. (2012). Beliefs about teaching science: The relationship between elementary teachers' participation in professional development and student achievement. *International Journal of Science Education, 34*(2), 153–166.

McDermott, L. C., Heron, P. R. L., Shaffer, P. S., & Stetzer, M. R. (2006). Improving the preparation of K-12 teachers through physics education research. *American Journal of Physics, 74*(9), 763–767.

McDermott, L. C., Shaffer, P. S., & Constantinou, C. P. (2000). Preparing teachers to teach physics and physical science by inquiry. *Physics Education, 35*(6), 411–416. Retrieved from http://www.phys.washington.edu/groups/peg/pubsa.html - abs2a

Mead, Louise S., & Mates, Anton. (2009). Why Science Standards are Important to a Strong Science Curriculum and How States Measure Up. *Evolution: Education and Outreach*, *2*(3), 359–371. doi:10.1007/s12052-009-0155-y

Moore, M. (2001). Teaching Evolution Do State Science Standards Matter? *Reports of the National Centre for Science Education*, *21*(1-2), 9–21.

Moseley, C., & Ramsey, S. J. (2008). Elementary teachers' progressive understanding of inquiry through the process of reflection. *School Science and Mathematics*, *108*, 49–57. doi:10.1111/j.1949-8594.2008.tb17804.x

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*(3), 315–340.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning / committee on development of an addendum to the national science education standards on scientific inquiry, center for science, mathematics, and engineering education*. Washington, DC: National Academy Press.

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, *2*. Retrieved from http://PAREonline.net/getvn.asp?v=8&n=2

Polly, D., McGee, J., Wang, C., Martin, C., Lambert, R., & Pugalee, D. K. (2015). Linking professional development, teacher outcomes, and student achievement: The case of a learner-centred mathematics program for elementary school teachers. *International Journal of Educational Research*, *72*, 26–37.

Texas Education Agency-Lonestar. (2013). Retrieved on http://loving1.tea.state.tx.us/lonestar/Home.aspx.

Texas Education Agency. (2011). Retrieved on http://ritter.tea.state.tx.us/perfreport/aeis/2011/.

Texas Education Agency. (2013). Retrieved on http://www.tea.state.tx.us/index.aspx.

Silverstein, S. C., Dubner, J., Miller, J., Glied, S., & Loike, J. D. (2009). Teachers' participation in research programs improves their students' achievement in science. *Science*, *326*, 440–442.

Southerland, S. A., Smith, L. K., Sowell, S. P., & Kittleson, J. M. (2007). Chapter 2 resisting unlearning: Understanding science education's response to the United States's national accountability movement". *Review of Research in Education*, *31*, 45–77.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, *30*, 23–28.

Supovitz, J. A. (2001). Translating teaching practice into improved student achievement. In S. H. Fuhrman (Ed.), *From the capitol to the classroom. Standards-based reform in the states* (pp. 81–98). Chicago: Univ. of Chicago Press.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, *37*(9), 963–980.

Tai, R. H. (2008). *Research on student interest and performance: Factors within a teacher's influence*. Washington, DC: Chemical Sciences Roundtable, National Academy of Science.

Tai, R., Loehr, J., & Sadler, P. (2005). Factors influencing success in introductory college chemistry. *Journal of Research in Science Teaching*, *42*(9), 987–1012. doi:10.1002/tea.20082

Thoemmes, F., & Kim, E. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*, 90–118.

van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. *Journal of Research in Science Teaching*, *38*, 137–158. doi:10.1002/1098-2736(200102)38:2<137::AID-TEA1001>3.0.CO;2-U

Van Overschelde, J. P. (2013). Project lead the way students more prepared for higher education. *American Journal of Engineering Education*, *4*(1), 1.

Vasquez, J. (2005, January 10). Bring back elementary science education. *National Science Teachers Association Express*. Retrieved from http://science.nsta.org/nstaexpress/nstaexpress_2004_01_10_letter.htm

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, *37*(8), 469–479.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad. Technical Report.*. National Staff Development Council.

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Educational Testing Service.

Wenglinsky, H., & Silverstein, S. C. (2007). The science training teachers need. *Educational Leadership*, *64*(4), 24–29.

Yoon, K. W., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (REL 2007- No. 033). REL Southwest.